DOCUMENT RESUME

ED 455 299 TM 033 156

AUTHOR Hafner, Anne L.
TITLE Evaluating the Impact of Test Accommodations on Test Scores of LEP Students & Non-LEP Students.
PUB DATE 2001-04-00
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).
PUB TYPE Numerical/Quantitative Data (110) -- Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; Analysis of Variance; *Elementary School Students; English (Second Language); Intermediate Grades; *Junior High School Students; Junior High Schools; *Limited English Speaking; Mathematics Achievement; Reading Achievement; *Scores; Test Results; Timed Tests
IDENTIFIERS *Testing Accommodations (Limited Engl Proficiency)

ABSTRACT

Using a quasi-experimental analysis of variance (ANOVA) design, this project examined the effects of the use of accommodations with students of limited English proficiency (LEP) and non-LEP students and whether the use of accommodations affected the validity of test score interpretations. Major accommodations examined were extra time, and extra time with extended oral presentation. Samples of fourth (n=292) and seventh grade students (n=159) were tested using the Terranova multiple assessment mathematics test, as well as a math skills test and the LAS reading comprehension test. Descriptive findings showed that LEP students scored lower than non-LEP students on mathematics tests and teacher-reported skill levels. Major predictors of mathematics achievement were LAS reading proficiency level (a proxy of LEP status), whether students received an accommodation, and teacher rating of reading skill. ANOVA analyses were conducted to compare the mean scores of students in accommodated tests versus those with no accommodation. These showed the accommodation effect was significant, with those students in the extra time condition showing the highest scores. A discriminant analysis showed that the best predictors of membership in different English fluency groups were Spanish fluency, time in the United States, reading grade, and mathematics test score. LEP students were more likely than non-LEP students to be misclassified into a fluency group. Examination of students' writing samples showed clear differences in mathematics and language achievement, depending on the student's language proficiency level. The project's results provide information on whether accommodations provide valid inferences for special needs children as well as non-LEP children. (Contains 1 figure, 16 tables, and 25 references.) (SLD)

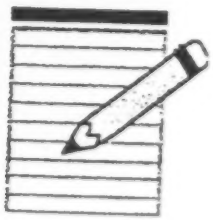
Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 455 299

Evaluating the Impact of Test Accommodations on Test Scores of LEP Students & Non-LEP Students

Anne L. Hafner
California State University, Los Angeles
5151 State University Drive
Los Angeles CA 90032
(323) 343-4361, ahafner@calstatela.edu

TM033156



PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

A. Hafner

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper presented at the annual meeting of the American Educational Research Assn, Seattle, WA, April 2001

2 BEST COPY AVAILABLE

Abstract

Using a quasi-experimental ANOVA design, this project examined the effects of the use of accommodations with LEP students, and non-LEP students and whether the use of accommodations affected the validity of test score interpretations. Major accommodations examined were extra time, and extra time with extended oral presentation. Samples of 4th and 7th grade students were tested using the Terranova multiple assessment math test, as well as a math skills test and the LAS reading comprehension test. Descriptive findings showed that LEP students scored lower than non-LEP students on math tests and teacher reported skill levels. Major predictors of math achievement were LAS reading proficiency level (a proxy of LEP status), whether students received an accommodation, and teacher rating of reading skill. ANOVA analyses were conducted to compare the mean scores of students in accommodated tests vs. those with no accommodation. These showed the accommodation effect was significant, with those students in the extra time condition showing the highest scores. A discriminant analysis showed that the best predictors of membership in different English fluency groups were Spanish fluency, time in US, reading grade and math test score. LEP students were more likely than non-LEP students to be misclassified into a fluency group. Examination of students' writing samples showed clear differences in mathematics and language achievement, depending on the student's language proficiency level. The project's results provide information on whether accommodations provide valid inferences for special needs children, as well as non-LEP children.

Acknowledgments

The author would like to express appreciation to the Delaware Dept. of Education, State of Delaware for funds received. Funds were initially received from the US. Department of Education. These funds enabled me to carry out the study. Publication of this report and the views expressed herein should not be construed as an endorsement of the views by the U.S Department of Education or the State of Delaware.
I would particularly like to thank Nancy Maihoff, from the state of Delaware, who has been so supportive throughout the process and Liru Zhang, who was very helpful.

In addition, the research would not have been possible without the guidance of a number of individuals on the California State University, Los Angeles team. They include the following:
I really appreciate the assistance give me by Kyoko Ito at CTB/McGraw Hill. I could not have done it without her gracious help.

Jamal Abedi, UCLA
Fery Hejri, ARDAC
Sharon Ulanoff, CSU San Marcos
Kyoko Ito, CTB/McGraw Hill
Cheryl Gilera, CSU Los Angeles
Terry Ray, CSU Los Angeles
Tracy Lee, CSU Los Angeles
Curt Mearns, Albuquerque Public Schools
Participating Teachers and Coordinators

Evaluating the Impact of Assessment Accommodations on Test Scores of LEP Students and Non-LEP Students

Table of Contents

Acknowledgments

I. Introduction and Objectives

II. Review of the Literature

III. Methods
- A. Sampling and subjects
- B. Independent and dependent variables
- C. Design and procedures
- D. Data analysis procedures

IV. Findings
- A. Descriptive findings
  - 1. Demographics and educational status
  - 2. Correlations
  - 3. Accommodation groups and random assignment
- B. Outcome Data
  - 1. Effects of accommodations: 4th grade students
  - 2. Regression analyses: predictors of math achievement
  - 3. Discriminant analysis: predictors of group membership
  - 4. Other analyses

V. Conclusions and Recommendations

References

I Introduction and Objectives

Appropriate inclusion of English learners (called LEP students in this proposal) and students with disabilities in large scale performance assessments is no small challenge, but the potential benefits are great. How can appropriate inclusion of such students in assessment programs contribute to the improvement of education outcomes students under reform? Cooley (1991) points out some of the possible uses of such assessments: they can inform policy, they can reform the curriculum and can increase accountability. However, there is no pat answer to the question of how best to 'appropriately' include LEP students and students with disabilities.

The educational reform initiatives under GOALS2000 and Improving America's Schools Act (IASA) call for assessment innovations in support of high standards to raise the achievement of *all students* including LEP students. NAEP has made strides in recent years in addressing students with disabilities and LEP student needs, including developing a side by side bilingual test, and allowing various accommodations. In addition, IDEA legislation mandates that states test students with disabilities.

In another recent development, federal legislation in 1997 regarding development of a Voluntary National Test (VNT) points to the importance of addressing the needs of disabled and LEP students, as well as inclusion and accommodation issues. Several such developments have converged to focus increased attention on the issue of including LEP students and students with disabilitiesand tracking these students' achievement and progress accurately. It is critical for state departments of education as well as local districts to be able to accurately assess and monitor the academic progress of all students with their testing programs.

The role of statewide assessment programs takes on increasing importance under education reform, as statewide tests become one of the primary measures of attainment of student performance standards. Educators are now looking to find ways to give LEP students and students with disabilities access to the full grade level appropriate curriculum and to carry out assessments that give these students the opportunity to show what they know and can do.

Although some agree that the move toward increased use of performance assessments may offer students a fairer and more contextualized method of ascertaining what they know and can do, others point out that new questions of validity arise. As assessment becomes increasingly embedded in instruction, it becomes more and more important for us to examine the validity of modifying or mediating assessments for various subgroups and to develop criteria/principles for the fair and valid administration of assessments to all students.

This project will contribute to the advancement of theory and knowledge in the area of the valid and fair assessment of all students. In light of the standards movement at the federal and state levels, the question is how can second language learners and students with disabilities be fairly held to the standards as well as included in assessments as much as possible? This project aims to help answer this question.

As mentioned in recent reports (see AIR, 1998a), very little research has been done on the use of accommodations with LEP students. NAEP has conducted some research but sample sizes in the 1996 administration were too small to evaluate the effects of accommodations on the technical characteristics of scores. NAEP did find that including scores for students who received accommodations did not have a significant effect on overall scale score results.

The purpose of this research study was to examine the effects of the use of accommodations with LEP students, and non-LEP students and whether the use of accommodations affects the validity of test score interpretations. If any of these are found to be "yes", we need to look at which accommodations affect test performance in which ways.

The study's research questions are the following:

- What are the effects of using specific accommodations on test scores of LEP students, LEP students with disabilities and non-LEP students?
- Do English proficient students benefit equally if allowed the same accommodations as LEP students?
- What accommodations provide valid inferences for LEP students and LEP students with disabilities?

II. Review of the Literature

In discussing the assessment of special needs students, one measurement concept that is important to consider is equivalence, which refers to the degree to which test scores can be used to make comparable (valid) inferences across diverse groups. A major concern is construct equivalence, or is a test measuring a construct (such as math knowledge) in a group that is equivalent to the construct being tested in other groups. In the case of LEP students, how can we determine whether a test measures a construct (math knowledge) only or whether English language proficiency is also being assessed? The same question goes for students with disabilities

In the field of performance assessment, few studies have focused on validating performance by different language groups. In addition, assessment administration has not been a focus of much work in the field; most of the interest has been on task development, scoring and general validation issues.

At the national level, the NAEP exams have allowed several types of accommodations on its exams, for students with disabilities and LEP students, depending on the exam and grade level. These include language accommodations such as a bilingual test book, bilingual dictionary or glossary; test setting accommodations such as one-on-one testing; extended time; read aloud or repetition of the test instructions; and accommodations for disabled children such as Braille, large print or computer equipment accommodations. (AIR, 1998a).

Many states (about 3/4) allow accommodations on at least one of their statewide assessments. The most frequently used are extra time (25 states), test setting accommodations (25-29 states); repeating directions (28 states), reading questions aloud (21 states), using word lists or dictionaries (14 states, translation of directions (19 states) and use of alternate assessments (11 states).(CCSSO, fall 1997)

Some test modifications and accommodations are unlikely to affect test scores and some are likely to affect scores. The 1985 Standards note that modifications of tests for individuals with handicapping conditions is, in general, desirable (Committee to Develop Standards for Educational and Psychological Tests, 1985). Much of the research in this area has only been done on individuals with disabilities. Little research has been done on the validity of test scores of LEP students vs. the scores of fully English proficient students. The Standards State there are few data to support conclusions about the effects of time modifications on test results (Committee to Develop Standards for Educational and Psychological Tests, 1985). A new revision of the Standards is underway, which provides more guidelines on assessing limited-English proficient students than in the previous version.

The research findings on the effects of giving extra time on essay exams are mixed. Many small or non-significant effects have been found, even with large differences in time allocation. Research does not prove that relaxing time limits significantly benefits any subgroup of examinees more than others, but the major subgroups used have been gender and ethnicity. A recent study (Powers and Fowles, 1996) found that additional time was equally beneficial to slow, average and fast test takers on an essay test (college students rated themselves as slow, medium or fast). In other words, the relative performance of slow, medium and fast test takers did not change much when more time was allowed (50% more time). Thus, these researchers found that the meaning of the test scores (construct validity) was unrelated to time limits. (Powers and Fowles, 1996, p. 448). Interestingly, students who said that English was not their best language were less likely to describe themselves as being able to write quickly.

In addition to extra time, other commonly used test accommodations include variations in presentation of test stimuli (e.g. simplifying words, reading aloud in English or L1, provision of a glossary); variations in response possibilities (oral vs. written response); and small group vs. individual administration.

Several accommodations were available for students in the NAEP 1996 math and science tests. In preliminary analyses of the data for comparing students tested with accommodations vs. those who did not have accommodations, NCES found little evidence of differential item functioning, although there were some statistical discrepancies. On the whole, including scores for students with disabilities and LEP students who received accommodations did not have a significant effect on the overall scale score results. However, these conclusions are said to be preliminary. (Mazzeo et al, 1998)

III Methods

This section will describe the methods of the study, including the subjects and sampling, the variables to be used, design and procedures, and data analysis procedures.

A. Sampling and subjects

School districts and school sites were recruited and selected, in the state of CA and New Mexico. Three groups of 4th graders were selected. The first group was made up of students identified as LEP either by an English language test or program placement. The second group was made up of non-LEP or English proficient students. The third group was made up of LEP students with disabilities as identified by an IEP or teacher designation. We originally hoped that a good number of LEP students with disabilities could be found and analyzed as a distinct group. Very little research has been done on this group. However, in our sample we did not find large numbers of these students. Only 22 students in the 4th grade were identified as being in a special education program. We were thus unable to analyze them separately as a unique group.

The sample selection process for the 4th grade sample was as follows. For the fourth grade, 4 schools with 4th grade classrooms were chosen purposively, matched by SES and size. The schools were low SES schools of average to large size, with a good proportion of LEP students. Within a school, at least 3 classes (at the 4th grade level) were chosen that contained some LEP students At the 4th grade level, assuming about 30 students per classroom, this would make a total of 15 fourth

grade classrooms for a sample size of about 450. With attrition and dropout, an effective sample size of about 430 was expected. As a result of attrition, eleven 4th grade teachers tested 292 fourth graders and seven 7th grade teachers tested 159 students.

At the 7th grade level, we wanted 15 classes: five schools with three classes each for about 450 students. With attrition and dropout, an effective sample size of only 160 was obtained. Classes sampled included general math courses made up of some English proficient and some LEP students.

Teachers were recruited from the chosen schools and trained in administering the tests and collecting survey data. Stipends were given to them for their work. A school site coordinator at each school coordinated the training workshop, materials distribution and mailing and was paid a stipend.

B. Independent and Dependent Variables

Accommodations chosen for this study were chosen because (a) many states including Delaware as well as districts report using them; (b) they appear likely to influence some student scores without changing the construct tested; and (c) they capture important aspects of test performance and comprehension. Accommodations that may influence or change the construct to be measured (such as reading the test items aloud to students) were not proposed here, as many believe that the use of such modifications indicates something different and usually results in scores being reported separately.

Accommodations that have been shown not to change the construct being tested were used in this study. They include:

(1) Timing: Extended time (ET). Extended time as an accommodation is widely used and has been shown to help improve scores on tests for some groups of examinees. LEP students may need more time to translate words or to comprehend the questions asked. Students with disabilities may need extra time because of learning disabilities or other disabilities. Based on the results of previous studies, the amount of extra time was 50% more time than standard.

(2) Extended Oral Presentation (EOP): Teachers were allowed to simplify test directions, re-read directions, provide additional examples, or read directions in students' native language. The activities that are permitted will be listed for the teachers. Extra time was also given with this accommodation, as teachers need additional time to do these activities.

In regard to dependent variables, the CTB/McGraw Hill Terranova math test was given in English to all of the students. This test is recommended, as it has known properties, validity, and reliability, contain multiple choice and constructed response items. A recent study used Terranova as well as a parallel test in Spanish in the 1998 administration (Supera), but very few students were actually getting most of their instruction in Spanish, and thus only 150 took the test in Spanish (Hafner, 1999) Also, the current policy interest seems to be in testing the students in English while giving accommodations, rather than providing the test in Spanish. The content area of math was chosen, as it is different from language arts, and is less language dependent than other subject areas.

Students in 4th and 7th grades were administered a short math basic skills test prior to the standardized math test, to obtain a measure of student ability or aptitude in math. They were also given the LAS reading test, which can be used either as a covariate or as a predictor of achievement.

Presentation to all students was made in terms of two half tests that were created using Terranova's Form A. The two half tests are roughly equivalent in terms of difficulty, content domain coverage, and both have multiple choice and constructed response items. Scores from each of the half tests can be reported on a common standardized scale. Terranova's multiple assessment includes selected response and constructed response items. The two item types can be scaled together. In addition, both norm referenced and curriculum-referenced scores (proficiency levels) can be produced.

C. Design and Procedures

The design is a quasi-experimental model. In this model, students participated in one form of the test accommodations (standard or accommodated). Classes were randomly assigned to one of the conditions. Table 1 shows the makeup off classes and student per grade and condition.

Table 1: Number of classes and students per condition

	4th Grade	7th Grade
Condition 1- Extra Time (ET)	5 (129)	3 (55)
Condition 2- Standard Administration	3 (83)	2 (41)
Condition 3- ET+Extended Oral Presentation	3 (80)	2 (63)
Total	11(292)	7 (159)

In the fourth grade the conditions were regular time, extra time and extended oral presentation (help with instructions). The three groups of students were LEP students, LEP students with disabilities and non-LEP students, thus a 3 x 2 factorial ANOVA design (see Table 2 for 4th grade design). For the seventh grade, the conditions were regular time, extra time, and extended oral presentation (see Table 3 for 7th grade design). Regular students and LEP students were included at 7th grade, for a 3 x 2 factorial design.

Table 2. ANOVA Design, 4th grade

	Regular time	Extra time	Extended oral	
Non-LEP students	60	60	60	
LEP students	40	40	40	
	100	100	100	N = 300

Table 3. ANOVA design, 7th grade

	Regular time	Extra time	Extended oral	
Non-LEP students	25	34	40	
LEP students	16	21	24	
	41	55	64	N = 160

At each site, a site coordinator and/or teacher administered the tests and collected the data on students and received a stipend for his/her work. The site coordinator or teachers were trained by

the PI. Teachers administered the tests, filedl out a survey for each child and collected other data on students. Training procedures were developed and carried out in the author's previous study in 1998, and were fine tuned for this administration.

In addition to the math test data, other student information were collected at the school site. To fully understand student performance, background and demographic information were collected via survey, along with other outcome variables to enable validation of the instrument. In 1998, an initial set of questions and data sources were piloted with teachers to try out items. Primarily, the variables include background variables (ethnicity, gender, primary language, self assessment of English proficiency, language classification status, years in school in the US, age, attendance record. SES) and educational variables such as grades, scores on norm referenced tests, primary language of instruction, and teacher ratings of student ability in math and reading. (see Appendix E for the list of variables)

At the university, pre-edit checks were conducted on the surveys. After data entry, post edit checks including consistency and range checks were performed to ensure quality data. Eventually, survey data were merged with CTB test data and analyzed.

D Data Analysis Procedures

To answer research question #3 (establishing validity for different groups of students), the validity of the constructs measured was ascertained by using correlations of test scores with other variables such as grades and teacher ratings. Research questions # 1 and 2 (on the effects of accommodations for different groups) will be answered by using a MANCOVA with math ability as covariate, testing for main effects for accommodation used and student subgroups.

CTB/McGraw Hill analysts developed the half tests, and have evaluated the quality of the half tests, including domain coverage, and difficulty, as well as their parallel structure. In addition to quantitative analyses, data from student writing samples were analyzed qualitatively for patterns and trends. To run analyses of variance and covariance, some of the variables had to re-coded. The following re-coding was performed. Due to the small number of students, accommodation variable CONDTN was re-coded from 3 categories (no accommodation and two forms of accommodation) to two categories (no accom/accommodation). This dummy variable was used in the multiple regression models. However, for analyses of variance and covariance, all three categories of accommodations were used. Number of years lived in US (TIMEUS) was re-coded to change the code for don't know from 0 to Missing, since 0 means no time in the US

IV. FINDINGS

The study's findings are presented as follows. First, descriptive findings such as demographics, educational status, correlations and accommodation groups by group assignment are presented. Next, outcome data findings are presented. These include the ANOVAs and ANCOVAs on the effects of accommodations, regression analyses, and discriminant analyses. Descriptive findings are presented here for the 4th and 7th grade students. CTB/McGraw Hill has not yet given us the full test data for the 7th graders, so the additional findings will be presented at a later date.

BEST COPY AVAILABLE

A. Descriptive Findings

1. Demographics and educational status

Tables 4a and 4b show the major demographics for the 4th and 7th grade students. About 60% of the 4th grade sample were Hispanic, 15% white, 20% African American, and 4% other. Most (83%) were in free lunch and a majority were in Title 1 (70%). About half (74) were in an LEP program. Only 16% or 22 were in a special education program and 41% were classified as LEP by LAS reading score.

Table 4a: Participant Demographics, 4th graders

Demographics	N, %
Male	120, 48%
Female	128, 52%
White	36, 15%
African American	48, 20%
Hispanic	149, 61%
Asian/Pacific Islander	4, 2%
American Indian/Native Alaskan/Other	6, 2%
Free lunch	192, 83%
Title I	99, 70%
LEP Program	74, 53%
Special Education Program	22, 16%
LEP (LAS Score)	82, 41%

Note: Number of students does not agree with totals on other tables because of missing data on one or more variables.

About 90% of the 7th grade sample were Hispanic, 3% white, 2% African American, and 5% other. Most (78%) were in free lunch and almost all were in Title 1 (94%). Only 8 students were in an LEP program. Only 19% or 18 students were in a special education program and 28% were classified as LEP by LAS reading score, fewer than at 4th grade.

Table 4b: Participant Demographics, 7th graders

Demographics	N, %
Male	61, 50%
Female	61, 50%
White	3,3%
African American	2, 2%
Hispanic	107, 91%
Asian/Pacific Islander	6, 5%
Free lunch	93, 78%
Title I	88, 94%
LEP Program	8, 9%
Special Education Program	18, 19%
LEP (LAS Score)	35, 28%

Note: Number of students does not agree with totals on other tables Because of missing data on one or more variables

Table 5 displays the means, standard deviations and numbers for demographic variables for the fourth graders. The scale for each variable is included in the table. The mean on English fluency is relatively high at 3.3, while the mean for Spanish fluency was low (2.5). Most of the students had lived in the US most of their lives.

Table 5: Educational Status, 4th graders

Variables*	N	Mean	St.Dev
Degree of English fluency (1-4)	248	3.31	.89
Degree of Spanish fluency (1-4)	240	2.50	1.34
No. of years lived in US (0-4)	172	3.76	.57
Reading grade (0-4)	190	2.59	.98
Math grade (0-4)	192	2.54	1.03
Reading skill (1-5)	220	3.08	1.06
Math skill (1-5)	220	3.06	1.02
LAS Standard score (2-100)	204	79.40	19.26
Math test score (0-20)	213	13.24	3.83
Yrs of Engl.instruction received (0-3)	197	2.86	.36
Received instruction in Sp/other (1-2)	180	1.77	.42
How long reading in English (0-4)	222	3.69	.72
Scale score 1 (Terrranova	219	608.84	52.46
Scale score 2 (Terranova)	219	615.56	45.89
Scale score total (Terranova)	219	1224.40	85.68
LEP status (1-3) (3=FEP)	204	1.60	.49

*See Appendix E Variable and Coding for scales.
Note: Number of students does not agree with totals on other tables because of missing data on one or more variables.

Table 6 shows the means, standard deviations, and numbers for the 7th grade students. As can be seen, 7th graders are similar to 4th graders on their means on most variables. The 7th graders scored slightly higher than 4th graders on degree of Spanish fluency and slightly lower on math grades and reading skill.

Table 6: Educational Status, 7th graders

Variables*	N	Mean	St.Dev
Degree of English fluency (1-4)	120	3.38	.83
Degree of Spanish fluency (1-4)	120	2.79	1.22
No. of years lived in US (0-4)	119	3.74	.63
Math grade (0-4)	104	1.90	1.37
Reading skill (1-5)	112	2.97	1.06
Math skill (1-5)	123	3.05	.97
LAS Standard score (2-100)	123	83.46	14.29
Math test score (0-20)	123	13.03	3.99
Yrs of Engl.instruction received (0-3)	119	2.91	.43
Received instruction in Spanish (1-2)	119	1.83	.44
Scale score 1 (Terrranova		N/A	N/A
Scale score 2 (Terranova)		N/A	N/A
Scale score total (Terranova)		N/A	N/A
LEP status (1-3) (3=FEP)		N/A	N/A

*See Appendix E Variable and Coding for scales.

NOTE: Number of students does not agree with totals on other tables because of Missing data on one or more variables.

2. Correlations

Table 7 shows the inter-correlations for all 4th graders among the Terranova math score, LAS reading test score, reading and math grades, math computation test score, degree of English and Spanish fluency, and years of English instruction. The tests show moderately high inter-correlations (between .64-.69). Reading and math grades are highly correlated at .83. English fluency is highly negatively correlated with level of Spanish fluency (-.82), indicating that a higher degree of English fluency is related to a low degree of Spanish fluency. Years of English instruction is moderately positively correlated with reading grade, math grade, English fluency and the LAS reading score (range is between .23- .52). Years of English instruction is negatively related to level of Spanish fluency (r=-.45**), thus those children with more years of English instruction have a lower level of Spanish fluency. Although there was a significant correlation between English fluency and Terranova math score and math test, there was no relationship between years of English instruction and math test scores.

Table 8 shows the correlations for 4th grade LEP students only. These correlations are somewhat lower than those for non-LEP students. For LEP students, there was a negative (but not significant) effect between English fluency and Terranova math and math test scores. The level of Spanish fluency was positively correlated with LAS Reading score, but this was not significant. As with the correlations for all students, years in English instruction were significantly correlated with reading grade (r=.32), English fluency (r=.45), and Spanish fluency (-.45). As contrasted with non-LEP students, LEP students did not show a significant correlation between English fluency level and LAS reading score, math test score, Terranova test score, reading and math grade. However, LAS reading score was significantly related to the tests and grades, for LEP and non LEP students. The English fluency rating, done by teachers, may reflect only oral fluency, which may not help the LEP students do well on tests.

Table 9 displays the correlations for non-LEP 4th grade students. These correlations are very similar to those from Table 7, for all 4th grade students.

Table 7. Correlations, all students, 4th grade

	TNmath	LAS-R	Mathtest	Readgrad	Mathgrad	EngFlu	SpanFlu	YrsEng
TN Math	1.0							
LAS-R	.64**	1.0						
Math test	.69**	.65**	1.0					
Readgrad	.59**	.69**	.42**	1.0				
Math grad	.63**	.68**	.50**	.83**	1.0			
English flu	.23**	.25**	.18**	.25**	.20**	1.0		
Spanfluenc	-.17*	-.06	-.06	-.09	-.07	-.82**	1.0	
Yrs Engli	.05	.23**	-.03	.33**	.24**	.52**	-.45**	1.0

* p<.05 ** p<.01

Table 8. Correlations LEP students, 4th grade

	TNmath	LAS-R	Mathtest	Readgrad	Mathgrad	EngFlu	SpanFlu	YrsEng
TN Math	1.0							
LAS-R	.56**	1.0						
Math test	.44**	.52**	1.0					
Readgrad	.31*	.49**	.22	1.0				
Math grad	.51**	.42**	.35*	.75**	1.0			
Engl fluen	-.07	-.11	-.10	.126	-.03	1.0		
Spanfluenc	.09	.21	.05	-.06	-.04	-.88**	1.0	
Yrs Engl	.01	-.088	-.19	.32*	-.00	.45**	-.45**	1.0

* p<.05 ** p<.01

Table 9. Correlations Non- LEP students, 4th grade

	TNmath	LAS-R	Mathtest	Readgrad	Mathgrad	EngFlu	SpanFlu	YrsEng
TN Math	1.0							
LAS-R	.52**	1.0						
Math test	.59**	.27**	1.0					
Readgrad	.62*	.57**	.31**	1.0				
Math grad	.61**	.52**	.47**	.76**	1.0			
Engl fluen	.14	.30**	.21*	.21*	.08	1.0		
Span fluen	-.23*	-.15	-.16	-.06	.001	-.82**	1.0	
Yrs Engl	.16	.17	-.05	.24*	..07	.55**	-.42**	1.0

* p<.05 ** p<.01

Table 10 shows the correlations for most of the 7th grade students. Patterns of inter-correlations are similar to those for all of the 4th graders.

Table 10. : Correlations, all students, 7th grade

	TNmath	LAS-R	Mathtest	Mathgrade	EngFlu	SpanFlu	YrsEng
TN Math	1.0						
LAS-R	.64**	1.0					
Math test	.62**	.50**	1.0				
Math grad	.55**	.28**	.60**	1.0			
English flu	.35**	.53**	.08	-.01	1.0		
Spanfluenc	-.21*	-.22*	-.06	.07	-.75**	1.0	
Yrs Engli	.19	.41**	.08	-.10	.45**	-.22*	1.0

* p<.05 ** p<.01

3. Accommodation Groups and Random Assignment

A cross tabulation was run to examine whether LEP students were randomly selected into different accommodation groups or classes. Although LEP students were more likely to be in the No accommodation condition, and non-LEPs were more likely to be in the Extra time + extended oral directions condition, the differences were not significant. See Table 11 for details.

Table 11: Accommodation group by LAS status

	LEP (LAS = 1 or 2)	Non-LEP(LAS = 3)	Total
Extra Time	36 43.9%	52 42.6%	88 43.1%
No Accommodation	28 34.1%	35 28.7%	63 30.9%
Extra Time + Extended Oral Directions	18 22.0%	35 28.7%	53 25.9%
Total	82 100.0%	122 100.0%	204

Note: Totals do not agree with those on other tables because of missing data

B. Outcome Data

1. Effects of Accommodations: 4th grade students

Table 12 shows the means and number of students by accommodation condition. As can be seen, the means are highest for the extra time condition. Table 13 shows the means on the two scale scores by LEP status. Students scoring at the LAS level 3 (fully English proficient) scored significantly higher than students scoring at level 1 on LAS (not English proficient) ($F=40.38$, $p<.000$, $F=4.31$, $p<.01$).

Table 12. Means & number of subjects by accommodation condition

Accommodation condition	**Scale score 1**	**Scale score 2**
No accommodation	x= 598.11, n = 64	x = 602.6 n = 64
Extra time	x = 619.24, n = 93	x = 634.32, n=93
Extra time + extra oral pres.	x= 604.32, n=61	x = 600.77, n=62
TOTAL	x = 608.85 n = 219	x = 615.56

Table 13. Means & number of subjects by LEP status

LEP status (LAS level)	**Scale score 1**	**Scale score 2**
LAS I (not Engl proficient)	x= 550.5, n = 17	x = 605.76 n = 17
LAS2 (LEP)	x = 587.04, n = 36	x = 616.1, n=35
LAS3 (FEP)	x= 637.02, n=90	x = 630.56, n=90
TOTAL	x = 614.15 n = 143	x = 623.97, n = 90

Results of ANOVA and ANCOVA

Total Terranova scale scores, SS1 and SS2, were used as dependent variables in two separate ANCOVA and ANOVA models, with accommodation code CONDTN serving as the independent variable. In the ANCOVA model, the computational math test score MATHTEST was used as a covariate, to control for math knowledge.

The results of ANCOVA analyses of Model 1 (SS1 as the dependent variable) showed no significant difference of the means by accommodations ($F=.29$, $df=2,151$, $p=.748$). However, the covariate was significant ($F=173.31$, $df=1,151$, $p=0.000$). But, the results of an ANOVA model (with no covariate) indicated that the accommodation effect was significant ($F=3.74$, $df=2,216$,

p=.033). (see Table 14)

The results of ANCOVA Model 2 (SS2 as the dependent variable) showed a significant accommodation effect (F=6.96, df=1,151,p=.001). When the effect of math test score was removed from the model as the covariate (ANOVA model), the effect of accommodation increased (F=15.31, df=2,216, p=0.000)(see Table 15). That is, the accommodation effect was significant even when a covariate was used. It must be indicated at this point that in the ANOVA and ANCOVA models, we lost a relatively large number of subjects due to missing data, (138 from the total of 292).

Table 14. Analysis of variance. SS1 as dependent variable

Source of Variation	SS	df	MS	F	Sig. F
Accommodation effect	18693.02	2	9346.51	3.74	.033*
Within (error)	581224.1	216	2690.85		
Total	599917.1	218	2751.91		

Table 15. Analysis of variance. SS2 as dependent variable

Source of Variation	SS	df	MS	F	Sig. F
Accommodation effect	56997.04	2	28498.5	15.31	.000**
Within (error)	402047.6	216	1861.33		
Total	459044.6	218	2105.71		

2. Regression Analyses

In order to conduct the regression, the accommodation code was re-coded as a dummy variable (0=no accommodation, 1=accommodation). This dummy variable was used as a predictor along with the following variables in the multiple regression:

ENGLFLU (English fluency)
LASLEVEL (LAS levels)
READSKIL (reading skills)
SPANFLU (Spanish fluency)
TIMEUS1 (Time lived in the us, recoded)

For all fourth graders, two regression models were created. The first model used scale score1 as the criterion variable and the accommodation dummy variable along with the other variables (listed above) were used as predictors. The second model used scalescore2 as the criterion variable with the same set of predictors.

Model 1 yielded an R square of .584 (over 58% of the variance of the criterion variable was explained by the predictors). Among the predictors, the effect of accommodation (t=3.19, p<001), LAS level (t=4.44, p<000) and reading skill (t=3.86, p<.00) were significant predictors in Model 1.

Model 2 yielded an R square of .308 (about 30% of the variance was explained). In this model, the accommodation variable (t=2.72, p<.01) and reading skill (t=2.72, p<.01) were significant predictors.

Separate regressions were run for LEP and non-LEP student groups. The two regressions run for LEP students showed R squared values of .28 (SS1) and .13 (SS2). Significant predictors for LEP children were reading skill ($t=3.1$, $p<.00$) in SS1 and a marginally significant time in US for SS2 ($t=1.85$, $p<.07$).

For non-LEP children, the regressions showed R squared values of .33 for SS1 and .33 for SS2. Significant predictors were receiving an accommodation ($t=4.1$, $p<.01$) and reading skill ($t=2.87$, $p<.01$) for SS1 and receiving an accommodation ($t=3.25$, $p<.002$) and reading skill ($t=2.76$, $p<.008$) for SS2. It is interesting to note that although LAS level (a proxy for LEP status) was significant predictor in the overall regression for all students, when the LEP and non-LEP student groups were run separately, results diverged. Only reading skill was a significant predictor for the LEP group.

3. Discriminant Analysis: Predictors of Group Membership

A discriminant analysis was done, using the following variables: Spanish fluency (SPANFLU), time in US (TIMEUS), LAS reading score (LASSTAN), math test score (MATHTEST), reading grade (READGRAD), and whether the student received instruction in Spanish (INSSPAN). This was done to ascertain which variables best predicted student placement in one of three categories by their teachers: limited English proficient, fluent in English as second language, or English as first language. Two significant functions were derived. The first one was dominated by the Spanish fluency variable. The second function was mainly made up of time in US, reading grade and math test score.

In looking at group means (see Table 16), we see that Group 2 (LEP) scored high on function 1 (Spanish fluency) and low on function two (time in US and reading ability). Group 3 (fluent in English as 2nd language) scored highest on function 2 (time in US and reading ability) and relatively high on function 1 (Spanish fluency). Group 4 (English as 1st language) scored the lowest on function 1 (Spanish fluency) and in between groups 2 and 3 on the second function (time in US and reading ability).

Table 16. Group means at functions

	FUNCTION 1	FUNCTION 2
Group 2: Limited Engl. proficient	2.29	-.93
Group 3: Fluent Engl. 2nd language	1.69	.95
Group 4: English 1st language	-2.69	-.08

Interestingly, group 3 is almost as high as group 2 on Spanish fluency, but outscores group 4 (English as first language) on function 2, which is made up of time in US, reading grade and math test score. This is a small but important group which should be examined (about 40 children). See Figure 1 following for the territorial map.

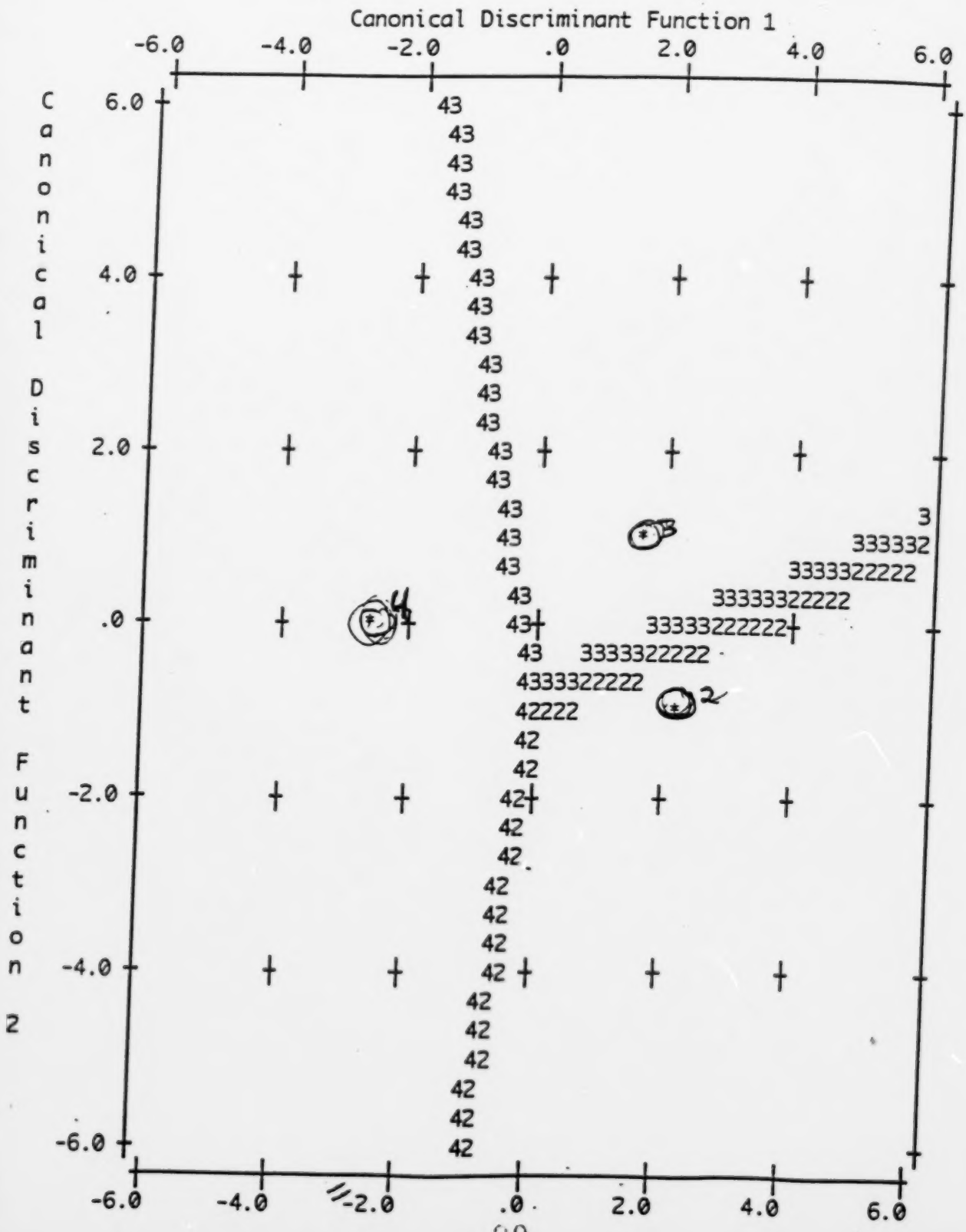
Insert Figure 1 about here

A classification analysis was also done in which subjects were grouped to see what percentage were correctly classified. While 87% of English as first language students and 83% of fluent in English as a 2nd language students were correctly classified, only 44% of LEP students (group 2) were correctly classified. 44 students were ungrouped originally, and 43 of these were classified in the fluent in English as second language group. The total percentage of students correctly classified was 74%.

Figiur e 1

Figure 1.

Territorial Map * indicates a group centroid
(Assuming all functions but the first two are zero)



4. Other Analyses

Writing samples that were embedded in the fourth grade exam were examined. The examination followed the principles of grounded theory (Strauss, 1987) focusing on data that pertain to the study of language and writing development. Several patterns emerged from this examination of the writing samples.

Examination of the writing samples clearly showed differences based on each student's language proficiency level. Students who were less proficient in English demonstrated very limited understanding of the math question that was asked, in other words, their limited language proficiency interfered with their understanding of the required task. These students often gave literal responses and/or simple translated the numbers into words (5 = five). They also had difficulty with even simple math tasks that required understanding of the written text to complete. This indicates that these less English proficient students had not yet developed their cognitive academic language proficiency or CALP (Cummins, 1999) and did not yet have enough English to understand the text and therefore the math task.

Along with that, many less proficient students demonstrated difficulty with both the math and the language. For example, it was common for the less proficient students to make errors on even simple questions both in terms of the math computations involved in addition to the aforementioned misreading of the required task. Furthermore, these students also showed a lack of familiarity with math words in English; some missed words as simple as subtraction (and performed a different math function).

While some students who had higher levels of English proficiency still had difficulty writing math responses, they were better able to match the task to what was required. For example, one student wrote, "I saw all the shapes, then in my mind I saw the shapes go on and that is how I got it" in response to a patterning question. Another student responded, "I counted the people" when asked how she estimated the number of people in the drawing.

The students who were more English proficient on the whole wrote more elaborate responses and used math language throughout their responses, e.g., estimation, patterns, etc. These students demonstrated a better understanding of math concepts as well as an understanding of math terminology. They also were able to describe the strategies they used to answer questions.

It is interesting to note that all students demonstrated accuracy in English spelling in their writing samples. Less English proficient students wrote shorter, more simplistic responses that did not always answer the questions posed, but were able to spell the words they chose conventionally.

V. Conclusions and Recommendations

The purpose of this study was to examine the effects of the use of accommodations with LEP and non LEP children, and whether the use of accommodations affects the validity of test score interpretations.

To answer our first research question, overall the use of accommodations did affect student test scores. In particular, students given the extra time accommodation showed higher mean scores. Regression analyses showed that receiving an accommodation did not significantly predict math achievement for LEP students, but did predict achievement for non LEP children.

The answer to the second question, do English proficient students benefit from accommodations? is a strong yes. In addition, non-LEP students showed a greater effect for accommodations.

Our third question, is not a straightforward one to answer. Which accommodations provide valid inferences for LEP students? LEP students show slightly lower correlations than non-LEP students. For LEP students, level of English fluency predicted only Spanish fluency (. -.88) and years of English instruction (.45**). For non-LEP children, English fluency level predicted LAS reading score, math test score, reading grade, as well as years of English instruction and Spanish fluency (negative).

LEP students were more likely than non-LEP students to have low test scores, grades and skills. Although LAS level (a proxy for LEP status in terms of reading) was a significant predictor in the overall regression for all students, for the group of LEP students, only teacher rating of reading skill significantly predicted Terranova math score. For non LEP students, reading skill and receiving an accommodation predicted math achievement. The Terranova math test seems to measure English reading proficiency in addition to math knowledge and skills. Extra time may enable students to translate words needed to solve problems. This may be especially true for word problems.

Shephard et al (1998) recommend that an accommodation should improve the performance of LEP students but should not improve the performance of English proficient students. In this study, both groups benefited. Thus, if extra time is offered by a state or local district, it should be offered to all students, not just LEP students.

Accommodations that allow students access to the test should be offered, as long as they do not influence the validity of the inferences made from them. Use of the extra time accommodation seems like a small price to pay to allow LEP students to show what they know and can do. However, non-LEP students should probably also be offered the extra time accommodation, in fairness.

Other findings include the fact that the LEP student group showed a great degree of heterogeneity, as evidenced by their large standard errors in the analyses. In addition, results of the discriminant analysis showed that LEP students were twice as likely than English as a first language students to be misclassified in the analysis.

More exploratory work needs to be done to examine the fluent in English as a second language group and to tease apart issues in the classification and identification of language fluency groups.

It should be kept in mind that this study should be considered an exploratory analysis. Because of small sample sizes in the cells of the design, there may be some confounding going on. We have learned that it is not easy to make generalizations about LEP students. It appears to be necessary to isolate unique accommodations for different subgroups ,as well as for non-minority groups.

In addition, it may be reasonable to move away from a research paradigm in which we make blanket generalizations about testing of all LEP students or about students in bilingual or ESL programs , and move toward an individual model.. Shepard et al (1998) note that very few LEP students receive accommodations specific to their language needs. Many schools and districts accommodate "all or none" of the LEP students or students with disabilities. Shepard et al (1998) suggest more training of school personnel so they can make better informed recommendations more targeted to the needs of individual English language learners.

REFERENCES

Abedi, J., Lord, C., and Plummer, J. (1995) *Language Background Report*. Los Angeles: UCLA Graduate School of Education, National Center for Research on Evaluation, Standards, and Student Testing.

AIR (1998a). Background paper reviewing laws and regulations, current practice, and research relevant to inclusion and accommodations for students with limited English proficiency. Palo Alto, CA author.

August, D., and McArthur, E. (1996) *Proceedings of the Conference on Inclusion Guidelines and Accommodations for Limited English Proficient Students in the National Assessment of Educational Progress* (December 5-6, 1994). Washington, DC: National Center for Education Statistics.

CCSSO (1992). Recommendations for improving the assessment and monitoring of LEP students. Washington DC: CCSSO report.

CCSSO (1997 fall). Annual Survey of State Student Assessment Programs. Washington, DC: author

Committee to Develop Standards for Educational and Psychological Tests of the American Psychological Association, American Educational Research Assn, (1985) *Standards for Education and Psychological Tests*: Washington DC: APA.

Council of Chief State School Officers, (1991). *Summary of State Practices Concerning the Assessment of and the Data Collection about LEP Students*. Washington, DC: Council of Chief State School Officers.

Gandara, P & Merino, B. (1993) Measuring the outcomes of LEP programs, test scores, exit rates and other mythological data. *Educational Evaluation and Policy Analysis, 165* (3), 320-338.

Garcia, E. & Pearson, P. (1994). Assessment and diversity. Review of Research of Education. (20) 337-391.

Hafner, A. (1995) Assessment Practices: Developing and Modifying Statewide Assessments for LEP Students. Paper presented at the annual conference on Large Scale Assessment sponsored by the Council of Chief State School Officers, June.

Hafner, A., Rivera, C., Vincent, C., and Lacelle-Peterson, M., (1995) Participation of LEP students in Statewide Assessment Programs. George Washington University: Unpublished report

LaCelle-Peterson, and Rivera, C. (1994). Is it real for all kids? A framework for equitable assessment policies for English language learners. Harvard Educational Review.

Lam, T & Gordon., W (1992, Winter) State policies for standardized achievement testing of limited English proficient students. *Educational Measurement: Issues and Practice*. 18-20.

Linn, Baker and Dunbar (1991). Complex performance based assessment: and validation criteria. Educational Researcher, 20, (8), 15-21.

(Mazzeo, J. et al. (1998). Increasing the participation of students w/disabilities and LEP students in NAEP: A special report on 1996 research activities. Washington DC: US Dept of Education.)

BEST COPY AVAILABLE

Mercado and Romero, M. (1993). Assessment of students in bilingual education. in M. B. Arias and U. Casanova (eds). Bilingual education: Politics, practice and research. Chicago: Univ. of Chicago Press.

Olson, J.F., and Goldstein, A.A. (1996). *Increasing the inclusion of Students with Disabilities and Limited English Proficient Students in NAEP.* NCES Focus on NAEP Series 2(1):1-5.

O'Malley, M & Valdez Pierce, L. (in press) Survey of state assessment practices affecting language minority students.

Proceedings of the Conference on Inclusion Guidelines and Accommodations for LEP Students. (1996). Washington, DC: NCES report.

Shepard, L. et al. (1998)

Strauss, A.L. (1987). Qualitative analysis for social scientists. New York: Columbia University Press.

Thurlow, M. (1995). National and state perspectives on performance assessment and students with disabilities. Washington DC: Council on Exceptional Children

US House. (1993) *Improving American's Schools Act.* Washington, DC: GPO: H.R. 6.

US Senate. (1993) *GOALS 2000: Educate American Act.* Washington DC: GPO, S. 1150.

Zehler, Hopstock, Fleischman and Greniuk (1994). An examination of assessment of LEP student. Arlington, VA: Development Associates.

BEST COPY AVAILABLE